匿名化语音中的帕金森病检测评估：性能与声学分析

Carlos Franzreb[1,2], *Francisco Teixeira*[2], *Ben Luks*[3,4], *Sebastian M¨oller*[3], *Alberto Abad*[2] ,[4]

[1] 德国人工智能研究中心，德国[2] INESC-ID，葡萄牙[3] 柏林工业大学，德国[4] 里斯本大学高等技术学院´，葡萄牙

carlos.franzreb@dfki.de

arXiv:2603.07544v1 [cs.SD] 8 Mar 2026

摘要

从语音中自动检测帕金森病（PD）是一种颇具前景的非侵入性诊断工具，但也引发了严重的隐私问题。说话人匿名化可以缓解这些风险，但可能会抑制帕金森病检测所需的病理信息。我们使用两个西班牙语数据集，评估两种匿名化器（STT-TTS 和 kNN-VC）在隐私与帕金森病检测之间的权衡。STT-TTS 能提供更好的隐私保护，但会抹除韵律信息，严重降低帕金森病检测性能。kNN-VC 能保留时长和基频轮廓等宏观韵律特征，其 F1 分数仅比原始基线低 3%–7%，这表明采用适当的匿名化方法时，保护隐私的帕金森病检测是可行的。最后，我们通过声学失真分析刻画了 kNN-VC 的具体弱点，为设计能够更好保留帕金森病信息的匿名化器提供了参考。

索引词：病理性语音、隐私、说话人匿名化

1. 引言

帕金森病（PD）是一种中枢神经系统的神经退行性疾病 [1], ，其患病率正在迅速上升 [2]。它会影响运动功能，导致震颤等运动症状，也会影响言语产生 [3]。口吃、构音含糊以及语速不受控制地加快，是帕金森病患者的常见症状。此前的研究发现，帕金森病患者存在构音困难 [4]。一项规模更大的研究 [5] 报告称，帕金森病患者与健康对照组在构音、嗓音、语速和流利度方面存在差异。

若能及早发现，帕金森病的治疗效果会更好，但疾病初期缺乏症状，使其难以检测。目前尚无用于识别帕金森病的确切实验室检测方法，医生要准确识别该病需要深厚的专业知识 [6]。这些困难推动了自动帕金森病检测研究 [7], , 包括基于语音的检测 [8, 9, 10], , 为侵入性更低、成本效益更高的自动帕金森病检测方法开辟了新途径。

除与帕金森病检测相关的标识特征外，语音还包含大量生物特征信息，由此引发隐私问题 [11]。面对法规要求，收集用于训练深度学习模型的数据十分困难；用户也可能拒绝使用会采集其身份信息的服务 [12]。说话人匿名化 [13] 为这一问题提供了解决方案：在保留语音效用的同时隐藏说话人身份。效用如何界定取决于使用场景 [14]；对于帕金森病语音，匿名化既可以隐藏病理特征以保护说话人身份，也可以保留这些特征，以便进行检测，同时隐藏其余敏感信息。后者更具挑战性，因为帕金森病信息与说话人身份彼此交织 [15]。

为了量化保留了多少帕金森病信息，我们利用现有的检测算法 [8]。我们考虑两种语音包含帕金森病患者的数据集（PC-GITA [16] 和Neurovoz[17]）以及两种匿名化方法：语音-文本-语音（STT-TTS）和kNN-VC [18]。我们还比较原始语音与匿名化语音的声学特征，以更好地了解哪些帕金森病信息会丢失，并估计为帕金森病患者提供的隐私增益。该分析的结果可为未来更好地保留帕金森病信息的匿名化器设计提供参考。复现我们实验所需的全部代码可在线获取[1]。

1.1. 相关工作

出于隐私保护、减少污名或确保公平使用语音驱动技术等考虑，若干研究采用匿名化来隐藏病理特征。针对口吃语音的研究比较了基于语音转换的匿名化与 STT–TTS [19],, 结果表明，基于文本的重新合成能够彻底去除口吃特征，在掩盖病理状况的同时实现强隐私保护。更广泛地说，先前研究指出，强匿名化或隐私保护变换可能会抑制临床相关的声学异常，从而可能掩盖神经系统或精神疾病 [20]。

相比之下，其他研究考察了匿名化后病理特征是否仍可识别，主要是为了评估匿名化语音在临床研究中的实用性。一项针对多种言语障碍的大规模评估 [21]表明，尽管匿名化会降低说话人验证的性能，但对于包括构音障碍和发声障碍在内的许多病症，下游病理相关任务（如障碍分类和可懂度评估）基本上仍能得到保留。此外，一项主观评估表明，声音嘶哑和震颤等病理相关声学属性在经过明确建模后能够在匿名化过程中保留下来，使匿名化语音仍具有临床参考价值 [22]。这些研究表明，匿名化不一定会消除病理标志；经过审慎设计，匿名化语音仍可用于病理识别和分析。然而，这两项研究均使用基于原始语音训练的说话人识别器评估隐私，而已有研究表明，这种做法会高估隐私保护效果 [23]。此外，它们并未探究匿名化如何在声学层面改变话语

尽管从匿名化语音中检测PD颇具前景，但也存在缺点，例如可能会丢失相关临床信息。还有一些不会降低声学信号质量的替代方法，例如密码学工具 [24], , 它们以保密方式处理信号；或者隐私感知特征提取方法 [25], , 它们在用户设备上运行，且仅共享临床相关信息。不过，前者计算成本高昂，而且需要复杂的配置才能

[1] https://github.com/carlosfranzreb/spane

确保隐私 [26], ; 而后者往往缺乏可解释性，且针对特定任务。相比之下，匿名化语音可以像原始语音一样进行处理，因此具有灵活性，也便于进行透明分析。

2. 方法

在本研究中，我们首次尝试评估常用匿名化器在帕金森病检测中的表现。我们提出将帕金森病检测作为一种新颖的匿名化器效用评估方法，用于评估能否从匿名化语音中识别出帕金森病，并探究帕金森病如何影响说话人的隐私。我们考虑了两个数据集：哥伦比亚西班牙语的 PC-GITA [16], 和卡斯蒂利亚西班牙语的 Neurovoz [17], 。我们聚焦于两项任务：句子，参与者需要说出一组固定的句子；以及独白，参与者自由发言约半分钟，内容可以是讲述日常生活（PC-GITA）或描述图像（Neurovoz）。为简洁起见，我们将 PC-GITA 称为 GT，将 Neurovoz 称为 NV，并在其后附加任务名称的首字母（例如，PC-GITA 句子 缩写为 GT-S）。GT-S 旨在评估说话人的韵律，说话人需要大声朗读语音平衡句。NV-S 包含常用谚语；为兼顾视力障碍者，说话人听取这些谚语，而非自行朗读。

我们考察了两种匿名化方法：语音转文本再转语音（STT-TTS）和 kNN-VC [18]。STT-TTS 由 Whisper-large [27] 自动语音识别和 Kokoro[2] 文本转语音串联而成：先将输入语音转写成文本，再进行合成，从而去除所有说话人信息 [28]。kNN-VC [18] 的隐私保护效果较差，但对源语音的保真度较高，因为它保留了韵律信息 [29, 30]。这是一种任意到任意的语音转换器：使用 WavLM [31] 提取特征，再根据余弦相似度，用四个最近邻目标特征的平均值替换这些特征，最后通过神经声码器（HiFiGAN [32]）进行合成。

kNN-VC 需要目标说话人的语音才能进行转换。我们使用众包哥伦比亚语（CC）数据集[33], , 其中包含 17 位说话人，每位平均录制 27 分钟。使用原始语音训练的帕金森病检测器进行初步实验后发现，使用 CC 数据集进行同性别选择，效果优于跨性别选择或不受限制的选择。即使保留说话人的状态（健康对照或帕金森病），使用 PC-GITA 目标说话人的效果也没有更好。因此，我们在 kNN-VC 中使用 CC 数据集进行同性别选择；对于 STT-TTS，也采用相同的选择方式，并使用其两位可用目标说话人（每种性别各一位）。每条话语都会随机选择一位目标说话人。

3. 效用评估

我们采用的 PD 评估基于自动帕金森病检测器（PDD） [8], , 该检测器使用 Wav2vec2 特征训练分类器 [34]。PDD 针对每项任务单独训练和评估，采用 5 折交叉验证；每个实验使用不同的随机种子重复 5 次。由于数据稀缺，PDD 无法很好地泛化到其他数据集或任务，但在多种语言的若干数据集上表现出色。因此，我们也对每项任务单独进行评估，并考虑以下两种评估场景：

1. **PDD-O：** PDD 使用原始语音训练。2. **PDD-A：** PDD 使用由待评估的同一匿名化器匿名化的语音训练。

PDD-O 代表最简单的使用场景：有人希望将自己的语音发送给 PDD，同时避免相关的隐私风险。PDD-A 则要求服务提供商使用匿名化数据训练 PDD，并且只能与特定的匿名化器配合使用。对于本研究而言，PDD-O 用于评估 PDD 和匿名化器的泛化能力，而 PDD-A 则量化匿名化语音中包含多少帕金森病信息。

表 1：不同保留性别特征的匿名化器、数据集和任务对应的 *PDD F1* 分数。*O* 和 *A* 分别表示 *PDD* 的训练数据为原始数据或匿名化数据。原始数据的 *PDD-A* 结果使用 *kNN-VC* 的 *PDD-A* 计算。词错误率仅使用 *Whisper-large* 针对句子任务（*S*）计算，因为独白任务（*M*）没有真实标注。

数据集	匿名	PDD-O	PDD-A	WER
GT-S	None	79±1	78±1	13
	kNN-VC$_r$	47±1	80±1	18
	kNN-VC	54±1	76±0	21
	STT-TTS	49±0	53±1	12
GT-M	None	80±3	64±4	—
	kNN-VC$_r$	49±3	80±1	—
	kNN-VC	55±4	74±1	—
	STT-TTS	35±2	49±1	—
NV-S	None	86±0	81±1	8
	kNN-VC$_r$	81±2	86±1	12
	kNN-VC	70±2	82±1	17
	STT-TTS	35±0	64±0	12
NV-M	None	85±3	56±8	—
	kNN-VC$_r$	76±2	82±2	—
	kNN-VC	70±9	78±1	—
	STT-TTS	54±5	59±1	—

3.1. PDD 结果

表 1 展示了两个数据集和任务的 PDD F1分数。使用 Whisper-large 计算词错误率（WER），以确保匿名化器的输出具有可懂度；两个匿名化器和两个数据集均满足这一要求。平均而言，帕金森病患者的可懂度估计值最高可达健康对照组的三倍（例如，原始 GT-S 的 $\mathrm{WER}_{HC} = 10$ 和 $\mathrm{WER}_{PD} = 31$ ）。基线 PDD-O（即，使用原始语音训练和评估的 PDD）在 Neurovoz 上的 F1 分数高于 PC-GITA，但每个数据集在不同任务中的分数一致。这些分数与先前研究报告的结果相当 [8]。

PDD-A 列中使用原始语音的 F1 分数是指使用原始语音、并在 kNN-VC 上训练的 PDD 进行的评估。有趣的是，它在句子 任务上表现良好，这意味着在匿名化语音上训练的 PDD 可用于从原始语音中检测帕金森病。对于独白 任务，性能下降更为明显。

STT-TTS 作为用于帕金森病检测的匿名化器被证明无效，因为所需的韵律和声学信息在转写 [19]中被去除。值得注意的是，由于 Whisper-large 不保留口吃和犹豫，因此只能利用转写错误来区分健康对照与帕金森病患者。正如预期，STT-TTS 在两项 PC-GITA 任务上的 PDD-A F1 分数均较低（$< 54\%$）。对于 Neurovoz，PDD-A 分数更高（NV-M 为 59%，NV-S 为 64%）。更高的性能可能源于增加的训练数据（1,696 而非 1,000），这使得 PDD-A 能够定位反复出现的转写错误。Neurovoz 还包含两句对所有经 STT-TTS 匿名化的健康对照说话人均为 WER=0 的句子，而这在 PC-GITA 中并未出现。

kNN-VC 的 PDD-O 分数明显低于原始话语：PC-GITA 低 25%，Neurovoz 低 15%。Neurovoz 的 F1 分数仍然很高（均为 70%）。

[2] https://github.com/hexgrad/kokoro

表 2：原始语音和两种匿名化器隐私评估的内部等错误率（百分比）。英语数据集的评估在 *MLS-ES*微调之前进行。

	匿名化器					
	None		**kNN-VC**		**STT-TTS**	
数据集	HC	PD	HC	PD	HC	PD
LS	0.5	—	6.9	—	35.6	—
EdAcc	3.0	—	17.9	—	43.3	—
MLS-ES	0.2	—	2.4	—	26.0	—
GT-S	1.0	1.2	20.8	21.2	36.4	38.4
GT-M	0.9	1.5	11.8	10.2	28.3	30.3
NV-S	1.4	1.6	23.9	24.7	55.8	58.5
NV-M	1.8	0.0	13.4	16.6	49.3	51.2

与基线 PDD-O 相比，PDD-A 在句子任务中的 F1 分数几乎没有下降：PC-GITA 下降 3%，Neurovoz 下降 4%。对于匿名化器而言，独白任务更具挑战性，与基线的差异更大：PC-GITA 为 6%，Neurovoz 为 7%。句子 任务可能更容易，因为说话人说的是相同的句子，因此所有说话人都会出现相似的韵律和发音错误。这一假设也解释了 kNN-VC 的 PDD-A 对原始语音的泛化能力。

3.2. 分析 kNN 中转换步骤的影响 -VC

鉴于 kNN-VC 与 PDD 的特征提取器（WavLM 和 Wav2vec2）性质相似，帕金森病信息的损失很可能发生在最后两个步骤：转换步骤或声码器。为了研究各步骤造成的损失，我们运行了 kNN-VC 的重新合成版本（在表 1 中称为 kNN-VC$_r$ ），将 WavLM 特征直接传递给声码器，绕过转换步骤。PDD-A 的结果表明，帕金森病信息的损失发生在转换步骤：除 NV-M 外，kNN-VC$_r$ 在所有任务上的 F1 分数都与原始基线相同；在 NV-M 上，其表现仍比 kNN-VC 高 4%。

PDD-O 的结果因数据集而异。对于 Neurovoz，kNN-VC$_r$ 的分数明显优于 kNN-VC，NV-S 和 NV-M 分别高出 11% 和 6%，不过后者的标准差较大（9%）。kNN-VC 中源说话人与目标说话人使用不同口音（卡斯蒂利亚语和哥伦比亚语），这可能是其得分较低的原因。然而，对于 PC-GITA，kNN-VC$_r$ 在两项任务上的表现都不如 kNN-VC。与 Neurovoz 的情况一样，差异仅在句子任务上较大（7%）。这些结果与我们使用目标说话人进行的初步实验相符：kNN-VC$_r$的表现介于使用 CC 数据集目标说话人的 kNN-VC 与使用 PC-GITA 目标说话人的 kNN-VC 之间。性能下降可能源于 PC-GITA 与 LibriSpeech（声码器的训练数据）之间的声学失配。将 PC-GITA 的话语转换为 CC 目标说话人的语音，或可缓解这种失配。

4. 隐私评估

现在，我们研究帕金森病对隐私估计的影响。帕金森病患者可能更容易被识别，因为不流畅现象会带来身份泄露的新风险，但前提是说话人识别器能够利用这些现象。我们采用基于 VoicePrivacy 挑战赛 2024 协议 [36], 的隐私评估 [35]，通过说话人识别器（ ECAPA-TDNN [37]）模拟对匿名化语音的攻击。每种匿名化器都有专门的说话人识别器，该识别器使用其匿名化数据进行训练。鉴于两个帕金森病数据集规模较小，我们先使用 Librispeech *train-clean-360* 训练说话人识别器，然后再使用多语种 LibriSpeech 的西班牙语子集（MLS-ES） [38]对其进行微调。由于 MLS-ES 的说话人数较少，直接使用该数据集进行训练并不理想。MLS-ES 的数据分布严重不均衡；为此，我们每位说话人最多只使用 1,000 条话语进行训练。

在研究了如何配置评估数据集后[35], ，我们使用 MLS-ES 测试集中的 20 条试验语句和 20 条注册语句。对于帕金森病数据集，我们将所有可用话语随机、均匀地分为试验集和注册集；对于独白任务，这意味着将每位说话人仅有的一条长话语分成两半。表 2 展示了所有数据集和匿名化器的等错误率（EER）。对于帕金森病数据集，我们分别评估每个组（健康对照和帕金森病），计算组内等错误率[28], ；该指标假设攻击者知道说话人所属的组，并评估每位说话人与组内其他成员相比有多容易被识别。

原始语音的结果验证了隐私评估：识别器几乎可以验证 MLS-ES 中所有语音对。MLS-ES 的结果也清楚地显示，两种匿名化器提升了隐私保护水平。STT-TTS 的隐私保护效果不够理想（EER$<50\%$），因为 MLS-ES 中每位说话人说出的内容各不相同，因此存在语音差异，而说话人识别器可以捕捉到这些差异 [28]。

两个帕金森病数据集的结果显示，匿名化器带来的隐私提升相近。对于 kNN-VC，隐私估计始终高于原始语音。STT-TTS 的隐私估计要高得多，在 Neurovoz 任务中实现了近乎完美的匿名化（EER$\approx 50\%$）。评估能够识别出在两项 PC-GITA 任务中经 STT-TTS 匿名化的说话人，这表明转写文本中仍保留了一些说话人信息。有趣的是，这些发现与 PDD 结果形成对比：在部分 Neurovoz 转写文本中，帕金森病信息仍然存在。

除原始 NV-M 和 kNN-VC 的 GT-M 外，帕金森病患者在所有情况下的隐私估计都更高。尽管差异很小，说话人识别器显然难以表征帕金森病患者，这可以用 MLS-ES 中缺少病理性语音来解释。综合考虑三种匿名化场景，Neurovoz 的 EER 总体上高于 PC-GITA，这表明 PC-GITA 的说话人更具辨识度。这一结果也解释了为什么 PC-GITA 的 PDD F1分数低于 Neurovoz：较高的说话人变异性使得寻找与说话人无关的特征更具挑战性。就两项任务而言，句子 的 EER 始终高于独白；对于较长的话语，说话人识别器表现更好。

5. 失真分析

为了更好地了解 kNN-VC 保留了哪些帕金森病信息，我们比较了 GT-M 原始样本与匿名化样本的声学特征；由于 GT-M 的时长较长，其特征也更可靠。我们选取了 35 项 DisVoice 特征 [39]，这些特征与帕金森病检测 [8]尤为相关。它们包括韵律、声门、发声和发音特征。我们使用原始特征拟合了一个标准缩放器，并用它转换原始特征和匿名化特征。这种归一化便于直接比较这些特征。

为量化 kNN-VC 引入的声学失真程度，我们使用两种互补指标比较原始话语与匿名化话语：地球移动距离（EMD） [40] 和互信息（MI） [41]。EMD 量化将原始分布转换为匿名化分布所需的代价。MI 量化原始特征与匿名化特征之间共享的非线性信息量。

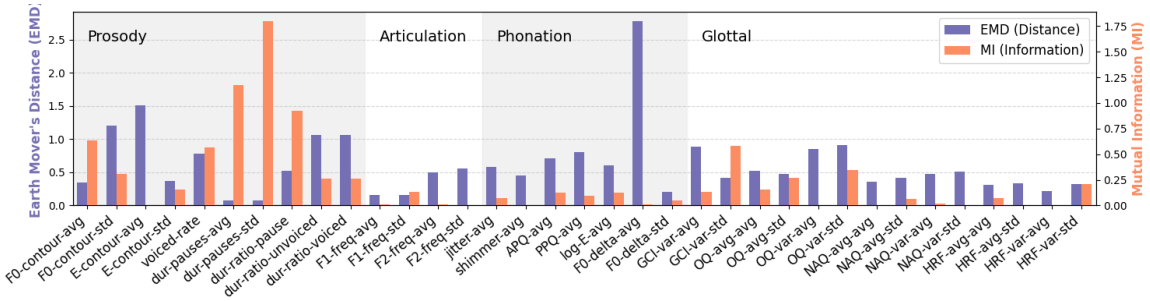


图1：*PC-GITA*独白任务中各特征的*EMD*与互信息 .

在我们的分析中，即使绝对声学值发生偏移，MI 也能揭示潜在的病理严重程度是否在匿名化过程中得以保留。就我们的 PDD 评估而言，MI 是 PDD-A 的代理指标，用于描述执行 PDD 所需的信息是否可用。另一方面，EMD 是 PDD-O 的代理指标，用于描述能否使用原始 PDD 对匿名化话语进行评估。

EMD与互信息的数值绘制于图1中。数据集中最严重的失真出现在发声类别内：基频导数的平均值呈现出最大的EMD（2.78），同时互信息接近于零（0.01）。匿名化语音的实际数值更低，这意味着音高轮廓更为平滑，推测其中不再包含帕金森病语音中出现的声带震颤和音高断裂，因为这些特征并不存在于健康目标语音中。构音和声门特征的互信息得分持续偏低（大多为 < 0.10）。匿名化后的构音和声门特征包含的是健康目标声道的信息，而非帕金森病患者的信息，从而使这些特征对匿名化后分类不再具有信息量。唯一的例外是声门闭合时刻（GCI）变异性的标准差（$MI = 0.58$），它是微扰的稳健度量。其较大的EMD值（0.25），以及平均GCI变异性保存较差，表明kNN-VC改变了GCI变异性的标准差，但各话语之间的差异仍然保留。

韵律是保留帕金森病信息最多的类别，但主要体现在宏观时间边界上。停顿时长的均值和标准差取得了最高的互信息值（1.18 和 1.80）以及较低的 EMD 值（< 0.08），这意味着它们未受 kNN-VC 改变，PDD-O 也应能够捕捉到这些信息。总体而言，kNN-VC 的逐帧映射成功保留了源说话人较大尺度的时长特征。相反，微观韵律成分遭到了扭曲。清音时长比发生了显著变化（EMD=1.06），支持了这样一种假设：kNN-VC 会抑制病理性的音节中微停顿和气声性无力。kNN-VC 也彻底改变了平均能量轮廓（EMD=1.51，MI=0.00）；匿名化语音的平均能量更高，这可能是健康目标语音所致。最后，平均基频轮廓也得到了中等程度的保留（EMD=0.35，MI=0.63），声门闭合脉冲波动亦是如此。单调音高或不规则呼吸组等病理特征，部分可用于从匿名化语音中检测帕金森病。PDD-A 可利用这两项特征从匿名化说话人中检测帕金森病，但 PDD-O 可能无法做到。

6. 讨论

在本研究中，我们全面分析了匿名化对从语音中检测帕金森病（PD）的影响及其潜力。结果表明，使用 PDD-O 检测经 kNN-VC 匿名化的语音中的 PD 不可行；为了使评估能够充分反映匿名化器保留的帕金森病信息，必须使用匿名化语音训练 PDD。有趣的是，在句子任务中，使用 kNN-VC 语音训练的 PDD-A 在评估原始语音时也同样有效，因为该任务的语音学特征受到限制，因此能够更有针对性地识别帕金森病信息。这意味着可以使用更容易收集的匿名化语音训练 PDD，而性能仅会略有下降。对于这项任务，PDD 无需学习能够泛化到其他句子的特征，因此韵律特征会变得更加重要，因为预计每个句子都遵循相似的韵律模式。尽管如此，同时使用原始语音和匿名化话语训练 PDD，可能会提高其稳健性，使其适用于这两类语音。另一方面，STT-TTS 未能保留帕金森病信息，这与先前的研究结果一致 [19]。

失真分析强调，kNN-VC 无法保留声门、构音和发声特征，声门闭合时刻（GCI）变异性除外，但其仍发生了偏移。只有韵律特征得到了较好保留，尤其是时长和平均基频轮廓。尽管如此，这些信息已足以使 PDD-A 的 F1 分数仅比原始基线低 3% 至 7%。这一结果表明，构音和发声特征对帕金森病检测并非至关重要，因为 kNN-VC 并未很好地保留它们。

重合成实验表明，这种帕金森病信息的损失几乎完全源于转换步骤，即源语音被健康目标语音所替换。可通过构建包含微颤和突然中断等表征帕金森病信息所需声学特征的目标说话人库来改进转换。k近邻算法也可能无法保留相关的声学信息。WavLM 特征主要由其音素内容主导 [42], , 目前尚不清楚，在考虑代表同一音素的目标特征时，k近邻是否会选择那些保留帕金森病信息的特征。同样不清楚 WavLM 捕获了多少帕金森病信息，PDD 的 Wav2vec2 也是如此；它足以实现最先进水平的帕金森病检测，但可能存在一些不足，例如我们发现在匿名化话语中被错误表征的声学特征。

关于隐私评估，说话人识别器未能充分利用帕金森病信息来识别说话人。在识别器的训练数据中加入病理性语音，应能提升其刻画帕金森病患者的能力。

7. 结论

据估计，STT-TTS 能为 Neurovoz 提供近乎完美的隐私保护，其对 PC-GITA 的隐私保护也远高于 kNN-VC。这种匿名化器非常适合需要将病理特征与说话者身份一同匿名化的使用场景。对于旨在检测帕金森病、同时保护说话者隐私的使用场景，kNN-VC 在 PDD 准确率与隐私保护之间取得了良好平衡。总体而言，这项研究进一步证明，匿名化是确保帕金森病检测隐私保护的一条有力途径。所开展的声学分析或可复制应用于其他

影响言语的疾病，从而有望拓展匿名化在影响言语的疾病检测中的使用场景。

8. 致谢

本研究部分获得德国联邦教育与研究部（BMBF 16KIS2048）的资助；同时获得葡萄牙国家资金的支持，该资金由科学与技术基金会（FCT）通过项目 UID/50021/2025（DOI：https://doi.org/10.54499/UID/50021/2025）和 UID/PRR/50021/2025（DOI：https://doi.org/10.54499/UID/PRR/50021/2025）提供；此外，还获得葡萄牙复苏与韧性计划、下一代欧盟资金在项目 C644865762-00000008（Accelerat.AI）下的资助。

9. 参考文献

[1] J. Parkinson，《关于震颤麻痹的论文，1817》，神经精神病学与临床神经科学杂志，第 14 卷，第 2 期，第 223–236 页，2002 年。[2] J. Zhu、Y. Cui、J. Zhang、R. Yan、D. Su、D. Zhao、A. Wang 和 T. Feng，《1980 至 2023 年帕金森病患病率的时间趋势：系统综述与荟萃分析》，《柳叶刀：健康长寿》，第 5 卷，第 7 期，第 464–479 页，2024 年。[3] A. A. Moustafa、S. Chakravarthy、J. R. Phillips、A. Gupta、S. Keri、B. Polner、M. J. Frank 和 M. Jahanshahi，《帕金森病的运动症状：统一框架》，神经科学与生物行为评论，第 68 卷，第 727–740 页，2016 年。[4] H. Ackermann 和 W. Ziegler，《帕金森性构音障碍中的构音缺陷：声学分析》，神经病学、神经外科学与精神病学杂志，第 54 卷，第 12 期，第 1093–1098 页，1991 年。[5] S. Skodda、W. Gr¨onheit、N. Mancinelli 和 U. Schlegel，《帕金森病病程中声音与言语障碍的进展：纵向研究》，帕金森病，第 2013 卷，文章编号 389195，2013 年。[6] B. Post、M. P. Merkus、R. M. A. de Bie、R. J. de Haan 和 J. D. Speel- man，《统一帕金森病评定量表运动检查：护士、神经科住院医师和运动障碍专科医生的评分是否可以互换？》运动障碍：运动障碍学会官方期刊，第 20 卷，第 12 期，第 1577–1584 页，2005 年。[7] G. Pahuja 和 T. N. Nagabhushan，《帕金森病检测现有机器学习方法的比较研究》，*IETE* 研究期刊，第 67 卷，第 1 期，第 4–14 页，2021 年。[8] D. Gimeno-G´omez、C. Botelho、A. Pompili、A. Abad 和 C.-D. Mart´ınez-Hinarejos，《揭示用于帕金森病诊断的自监督语音表示的可解释性》，*IEEE*信号处理精选主题期刊，第 1–14 页，2025 年。[9] C. Quan、K. Ren 和 Z. Luo，《一种利用语音动态特征检测帕金森病的深度学习方法》，*IEEE Access*，第 9 卷，第 10239–10 252 页，2021 年。[10] M. La Quatra、J. R. Orozco-Arroyave 和 M. S. Siniscalchi，《基于语音检测帕金森病的双语双头深度模型》，载于*ICASSP*，2025 年，第 1–5 页。[11] A. Nautsch、C. Jasserand、E. Kindt、M. Todisco、I. Trancoso 和 N. Evans，《《通用数据保护条例》与语音数据：法律与技术界的思考，以及迈向共同理解的第一步》，载于*Interspeech*。ISCA，2019 年，第 3695–3699 页。[12] C. Franzreb、A. Das、H. Gieseler、E. C. Jahn、T. Polzehl 和 S. M¨oller，《面向远程心理治疗的视听匿名化：主观评估》，载于第 *4* 届语音通信安全与隐私研讨会。ISCA，2024 年，第 102–110 页。[13] Q. Jin、A. R. Toth、T. Schultz 和 A. W. Black，《语音转换实现说话人去身份化》，载于*ICASSP*。IEEE，2009 年，第 3909–3912 页。[14] S. Meyer 和 N. T. Vu，《语音匿名化的使用场景》，载于第 *5* 届语音通信安全与隐私研讨会。ISCA，2025 年。

[15] P. Champion、D. Jouvet 和 A. Larcher，“构建说话人匿名化系统，只需解耦表示就够了吗？”，载于 Interspeech 2022。ISCA，2022 年，第 2793–2797 页。[16] J. R. Orozco-Arroyave、J. D. Arias-Londo˜no、J. F. Vargas-Bonilla、M. C. Gonz´alez-R´ativa 和 E. N¨oth，“用于分析帕金森病患者的全新西班牙语语音语料库数据库”，载于 LREC。欧洲语言资源协会（ELRA），2014 年，第 342–347 页。[17] J. Mendes-Laureano、J. A. G´omez-Garc´ıa、A. Guerrero-L´opez、E. Luque-Buzo、J. D. Arias-Londo˜no、F. J. Grandas-P´erez 和 J. I. Godino-Llorente，“NeuroVoz：卡斯蒂利亚西班牙语帕金森病患者语音语料库”，Scientific Data，第 11 卷，第 1 期，第 1367 页，2024 年。[18] M. Baas、B. van Niekerk 和 H. Kamper，“仅用最近邻进行语音转换”，载于 Interspeech。ISCA，2023 年，第 2053–2057 页。[19] J. Hintz、S. Bayerl、Y. Sinha、S. Ghosh、M. Schubert、S. Stober、K. Riedhammer 和 I. Siegert，“口吃语音匿名化——移除说话人信息，同时保留话语内容”，载于第三届语音通信安全与隐私研讨会。ISCA，2023 年，第 41–45 页。[20] C. Diaz-Asper、L. A. Bongo 和 B. Elvev˚ag，“在临床研究的语音匿名化中权衡个人隐私与数据效用”，npj Digital Medicine，第 8 卷，第 1 期，第 616 页，2025 年。[21] S. Tayebi Arasteh、T. Arias-Vergara、P. A. P´erez-Toro、T. Weise、K. Packh¨auser、M. Schuster、E. Noeth、A. Maier 和 S. H. Yang，“应对说话人匿名化面临的挑战，在保障病理性语音隐私的同时维持其效用”，Communications Medicine，第 4 卷，第 1 期，第 182 页，2024 年。[22] S. Ghosh、M. Jouaiti、A. Das、Y. Sinha、T. Polzehl、I. Siegert 和 S. Stober，“老年人及病理性语音匿名化：利用 DDSP 和基于示例的查询进行语音转换”，载于 Interspeech。ISCA，2024 年，第 4438–4442 页。[23] P. Champion，《语音匿名化：说话人匿名化技术的评估与设计》，博士论文，洛林大学，2023 年。[24] F. Teixeira、A. Abad 和 I. Trancoso，“保护隐私的副语言任务”，载于 ICASSP。IEEE，2019 年，第 6575–6579 页。[25] V. Ravi、J. Wang、J. Flint 和 A. Alwan，“通过说话人解耦提升基于语音的抑郁症检测的准确性与隐私保护”，Computer Speech & Language，第 86 卷，第 101605 页，2024 年。[26] F. Teixeira，《用于远程语音处理的隐私保护机器学习》，博士论文，里斯本大学高级技术学院，葡萄牙，2024 年。[27] A. Radford、J. W. Kim、T. Xu、G. Brockman、C. McLeavey 和 I. Sutskever，“通过大规模弱监督实现稳健的语音识别”，arXiv 预印本 arXiv:2212.04356，2022 年。[28] C. Franzreb、A. Das、T. Polzehl 和 S. M¨oller，“LibriSpeech 中的内容泄漏及其对说话人匿名化隐私评估的影响”，2026 年。[29] N. Gengembre、O. Le Blouch 和 C. Gendrot，“通过语音转换解耦韵律与音色嵌入”，载于 Interspeech。ISCA，2024 年，第 2765–2769 页。[30] C. Franzreb、A. Das、T. Polzehl 和 S. M¨oller，“Private kNN-VC：对转换语音进行可解释匿名化”，载于 Interspeech。ISCA，2025 年，第 3224–3228 页。[31] S. Chen、C. Wang、Z. Chen、Y. Wu、S. Liu、Z. Chen、J. Li、N. Kanda、T. Yoshioka、X. Xiao、J. Wu、L. Zhou、S. Ren、Y. Qian、Y. Qian、J. Wu、M. Zeng、X. Yu 和 F. Wei，“WavLM：面向全栈语音处理的大规模自监督预训练”，IEEE Journal of Selected Topics in Signal Processing，第 16 卷，第 6 期，第 1505–1518 页，2022 年。[32] J. Kong、J. Kim 和 J. Bae，“HiFi-GAN：用于高效、高保真语音合成的生成对抗网络”，载于 NeurIPS，第 33 卷。Curran Associates, Inc.，2020 年，第 17 022–17 033 页。[33] A. Guevara-Rukoz、I. Demirsahin、F. He、S.-H. C. Chu、S. Sarin、K. Pipatsrisawat、A. Gutkin、A. Butryna 和 O. Kjartansson，“为低资源文本转语音众包采集拉丁美洲西班牙语数据”，载于 LREC。欧洲语言资源协会（ELRA），2020 年，第 6504–6513 页。

[34] A. Baevski、Y. Zhou、A. Mohamed 和 M. Auli，“wav2vec 2.0：用于语音表示自监督学习的框架”，神经信息处理系统进展，第 33 卷，第 12 449–12 460 页，2020 年。[35] C. Franzreb、A. Das、T. Polzehl 和 S. M¨oller，“优化说话人匿名化器隐私评估所用的数据集”，载于第五届语音通信安全与隐私研讨会。ISCA，2025 年。[36] N. Tomashenko、X. Miao、P. Champion、S. Meyer、X. Wang、E. Vincent、M. Panariello、N. Evans、J. Yamagishi 和 M. Todisco，“VoicePrivacy
2024 2024, 2404 02677 挑战评估
[ ] [ ] 2404 02677
计划”，6 月。arXiv: . eess . Online . Available: http://arxiv.org/abs/ .[37] N. Dawalatabad、M. Ravanelli、F. Grondin、J. Thienpondt、B. De-splanques 和 H. Na，“用于说话人日志的 ECAPA-TDNN 嵌入”，载于*Interspeech*。ISCA，2021 年，第 3560–3564 页。[38] V. Pratap、Q. Xu、A. Sriram、G. Synnaeve 和 R. Collobert，“MLS：用于语音研究的大规模多语言数据集”，载于*Interspeech*。ISCA，2020 年，第 2757–2761 页。[39] J. V´asquez-Correa、J. Orozco-Arroyave、T. Bocklet 和 E. N¨oth，“朝着自动评估帕金森病患者构音障碍程度迈进”，沟通障碍期刊，第 76 卷，第 21–36 页，2018 年。[40] Y. Rubner、C. Tomasi 和 L. Guibas，“一种用于分布的度量及其在图像数据库中的应用”，载于第六届国际计算机视觉会议。IEEE，1998 年，第 59–66 页。[41] T. M. Cover 和 J. A. Thomas，信息论基础。Wiley，2001 年。[42] A. Sicherman 和 Y. Adi，“用于口语语言建模的离散自监督语音表示分析”，载于*ICASSP*。IEEE，2023 年，第 1–5 页。